# Machine Learning Approach in Particle Pair Investigation

Serpil Yalcin Kuzu*, Ayben Karasu Uysal
*Department of Physics, Firat University*

XIV. International Conference on Nuclear Structure Properties (NSP2021)
2-4 June, 2021

OUTLINE

- Motivation
  - Picture of Heavy Ion Collisions
  - Why dielectrons?
  - Why Machine Learning (ML)?
- Analysis:
  - Model: Random Forest Classifier
  - Experimental Setup
    - Data Set
    - Application of Model
- Results
  - ROC AUC Interpretation
  - Precision, Recall and F-1 Scores
- Conclusions

o **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

o **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

o **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores

o **Conclusions**
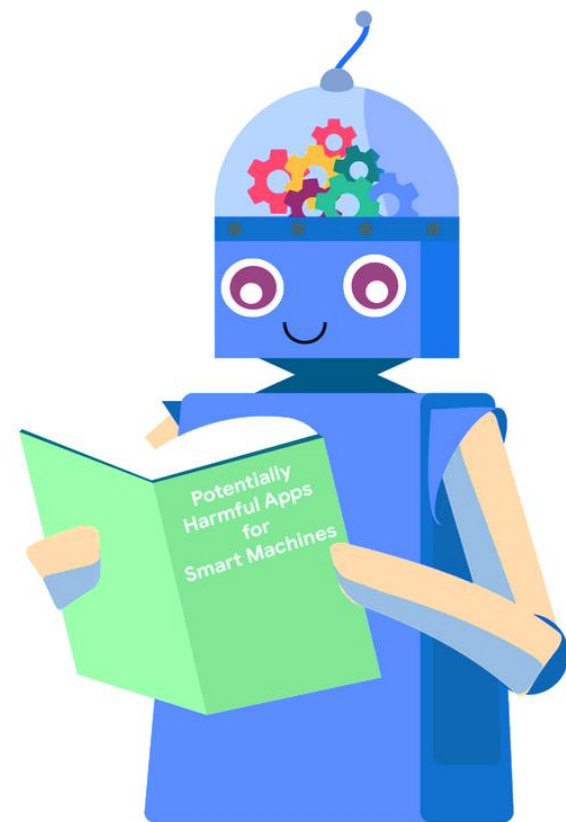
Heavy Ion Collisions

1. Initial state
- ions travelling with velocity near the speed of light

○ **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

○ **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

○ **Results**
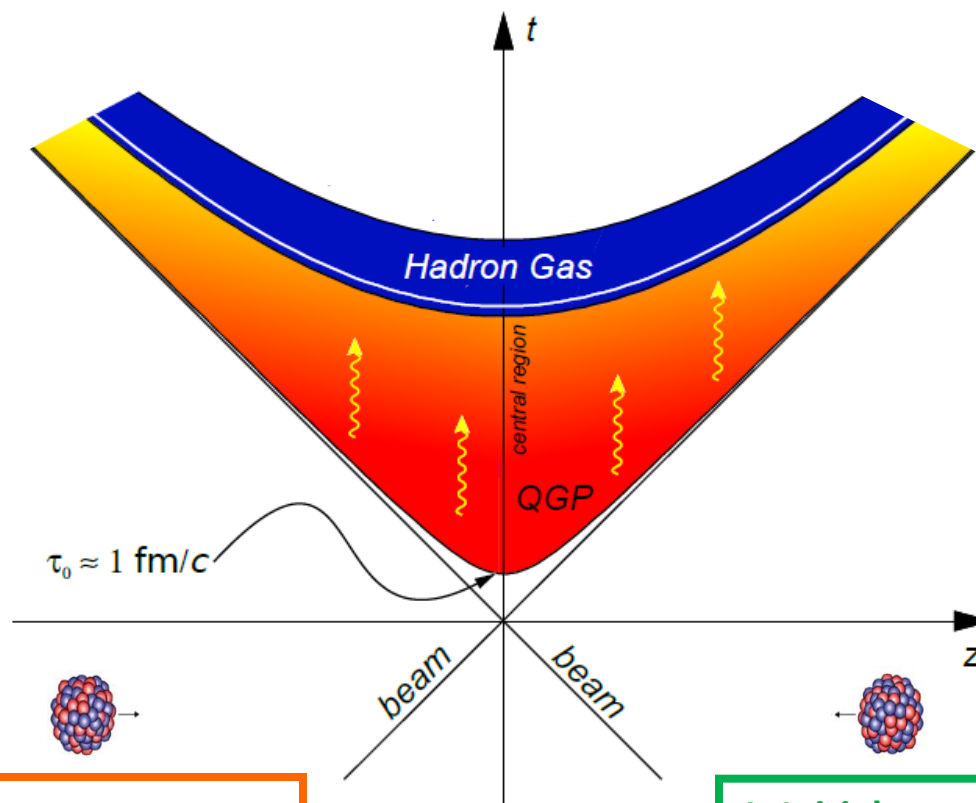 • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
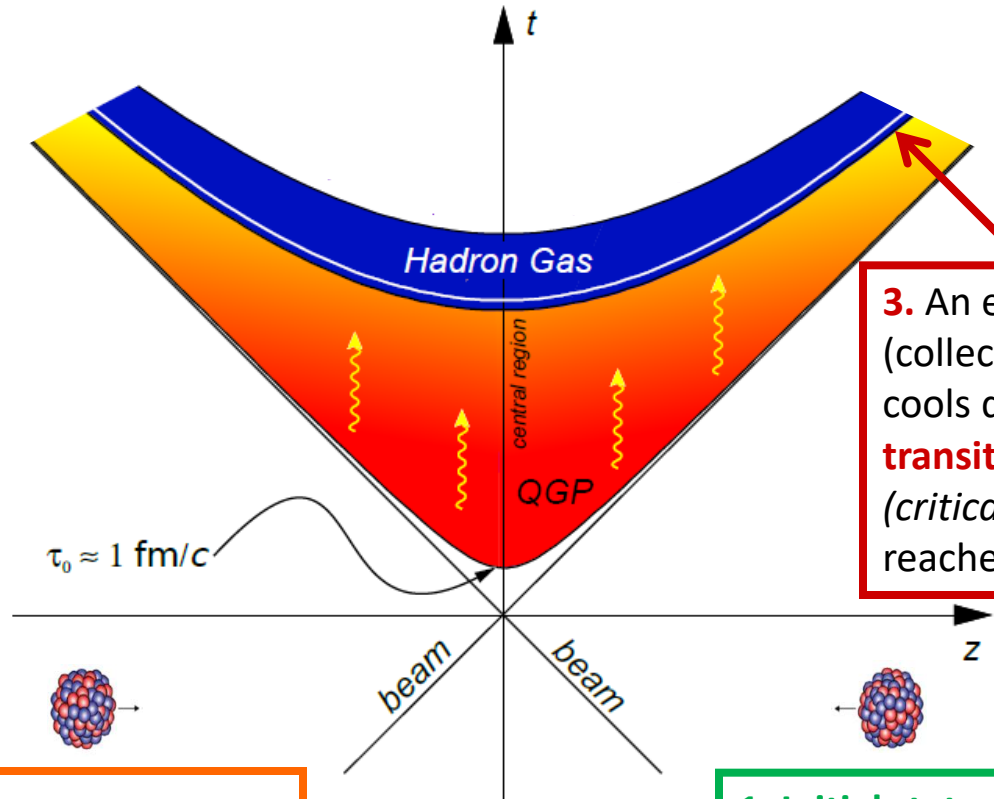○ **Conclusions**
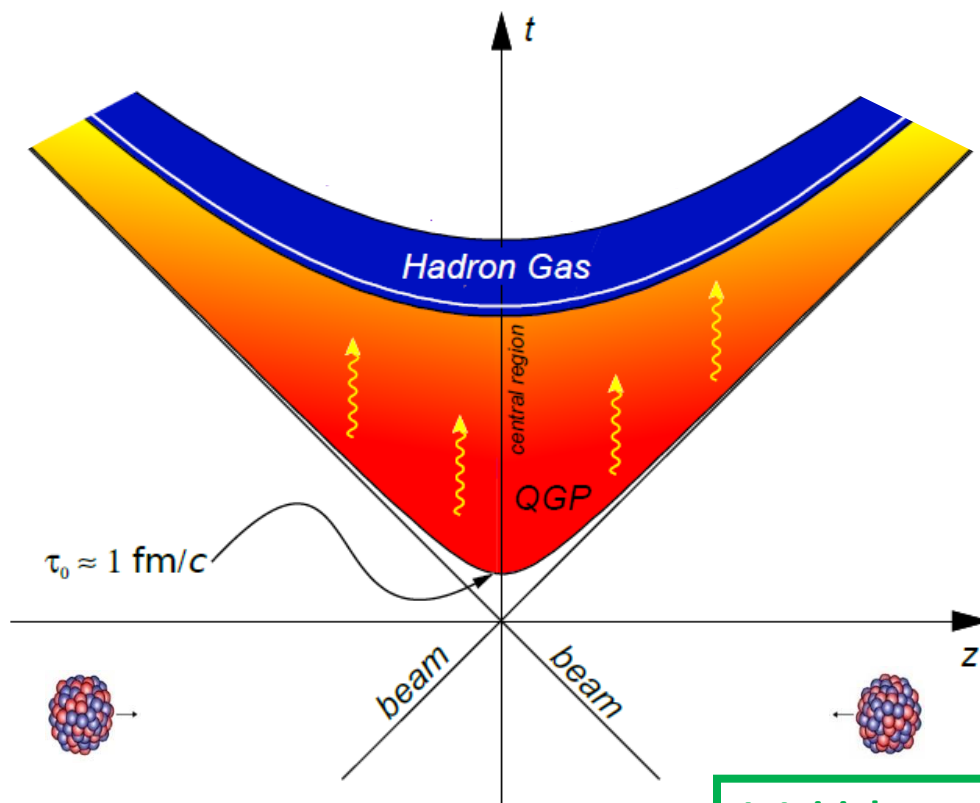
Heavy Ion Collisions

2. QGP formation
- pre-equilibrium and thermalization
$\tau_0 \approx 1$ fm/c

1. Initial state
- ions travelling with velocity near the speed of light

o **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

o **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

o **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores

o **Conclusions**

Heavy Ion Collisions

3. An expansion of the system (collective evolution) that cools down until the **phase-transition temperature T_c** *(critical temperature)* is reached.
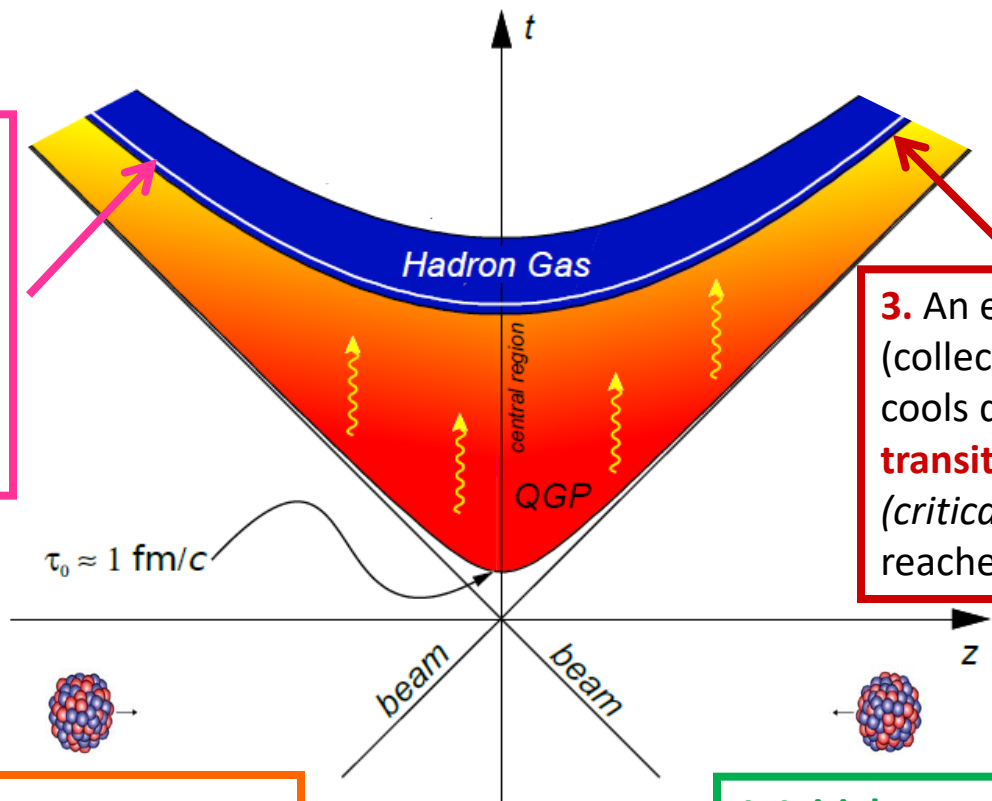
2. QGP formation
- pre-equilibrium and thermalization $\tau_0 \approx 1$ fm/c

1. Initial state
- ions travelling with velocity near the speed of light

o **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

o **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

o **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
o **Conclusions**

Heavy Ion Collisions

4. At T_c partons freeze out into hadrons. The hadron gas expands until inelastic collisions stop at the **chemical freeze-out** (T_{ch}).

3. An expansion of the system (collective evolution) that cools down until the **phase-transition temperature T_c** *(critical temperature)* is reached.
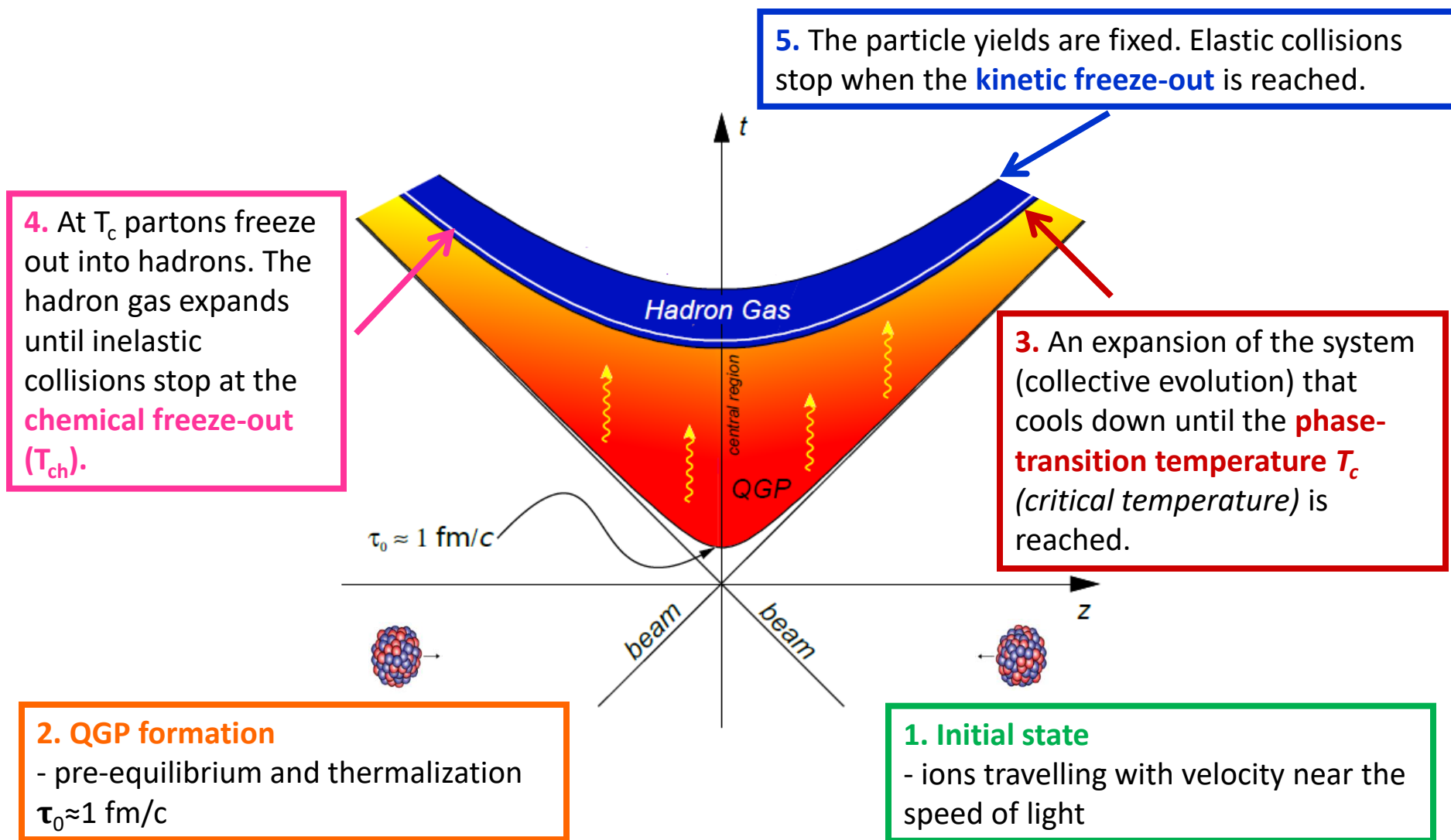
2. QGP formation
- pre-equilibrium and thermalization $\tau_0 \approx 1$ fm/c

1. Initial state
- ions travelling with velocity near the speed of light

○ **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

○ **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

○ **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores

○ **Conclusions**

Heavy Ion Collisions

5. The particle yields are fixed. Elastic collisions stop when the **kinetic freeze-out** is reached.

4. At T_c partons freeze out into hadrons. The hadron gas expands until inelastic collisions stop at the **chemical freeze-out** (T_{ch}).

3. An expansion of the system (collective evolution) that cools down until the **phase-transition temperature** T_c *(critical temperature)* is reached.

2. QGP formation
- pre-equilibrium and thermalization $\tau_0 \approx 1$ fm/c

1. Initial state
- ions travelling with velocity near the speed of light

o **Motivation**
- Heavy Ion Collisions
- Why Dielectrons? Why Machine Learning?

o **Analysis**
- Model: Random Forest Classifier
- Experimental Setup: Data/ Application of Model

o **Results**
- ROC AUC Interpretation/ Precision, Recall and F-1 Scores

o **Conclusions**

Heavy Ion Collisions

6. The particles leave the collision area and can be detected by detectors.

5. The particle yields are fixed. Elastic collisions stop when the **kinetic freeze-out** is reached.

4. At T_c partons freeze out into hadrons. The hadron gas expands until inelastic collisions stop at the **chemical freeze-out** (T_{ch}).

3. An expansion of the system (collective evolution) that cools down until the **phase-transition temperature T_c** *(critical temperature)* is reached.

2. QGP formation
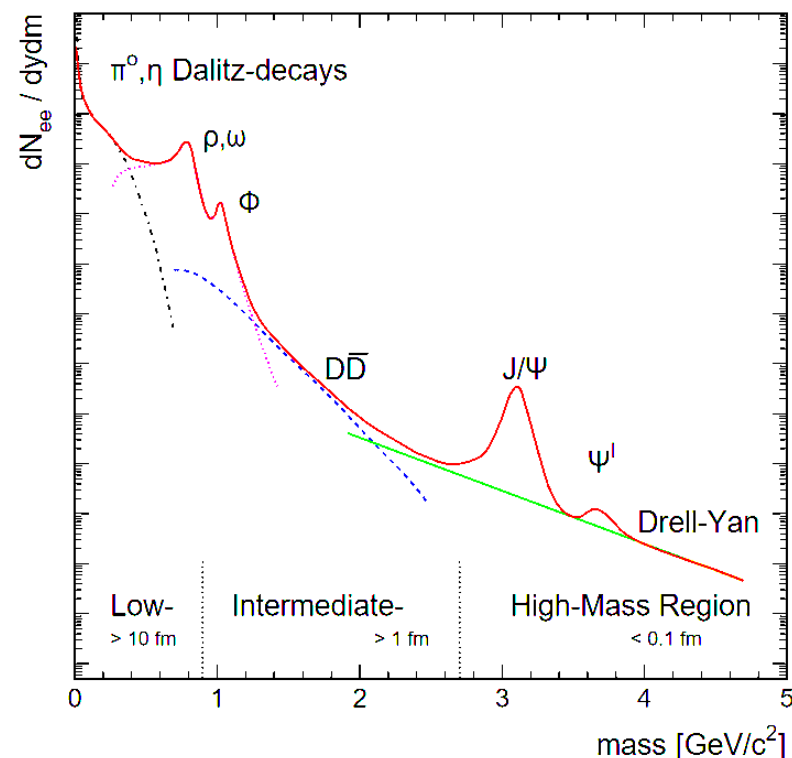- pre-equilibrium and thermalization $\tau_0 \approx 1$ fm/c

1. Initial state
- ions travelling with velocity near the speed of light

Freeze-Out

Hadron Gas

central region

QGP

$\tau_0 \approx 1$ fm/c

beam   beam

t

z
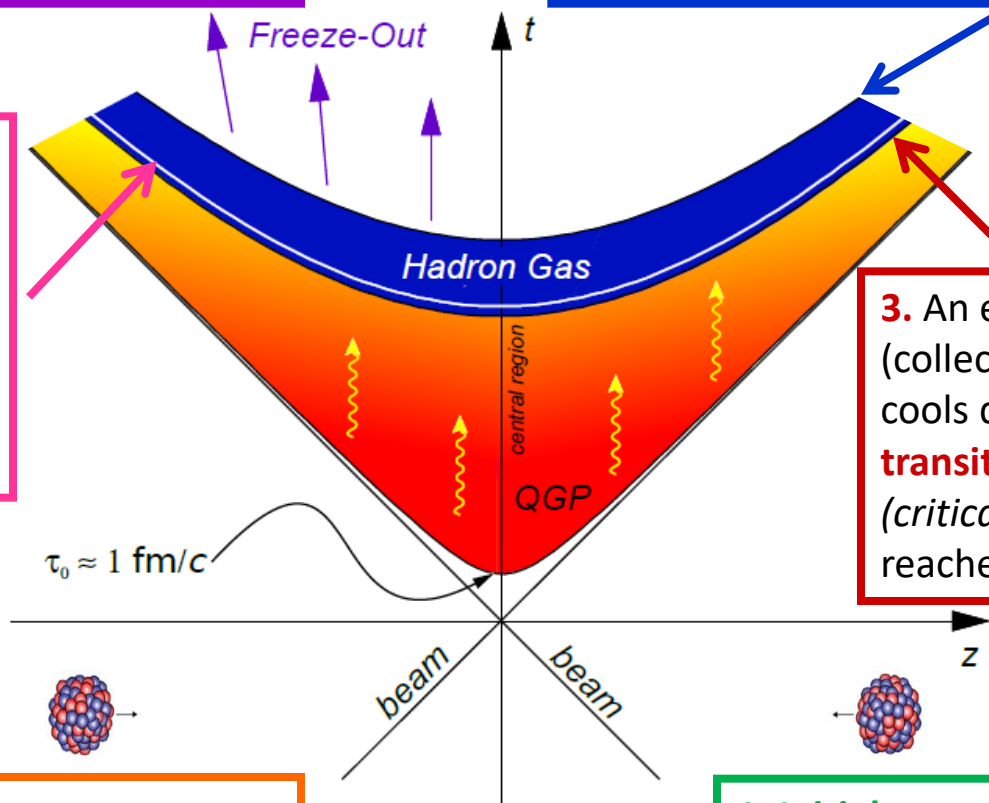
Why Dielectrons?

Electron-positron pairs *(dielectrons)* are the unique tools to study evolution of medium created in the collisions

- Due to lack of strong interactions they can be used to probe the inner regions of collisions
  - ○ no in medium effects on dielectrons
- They are produced at all stages of the collisions
  - ○ provide information about the whole space-time evolution of the system.

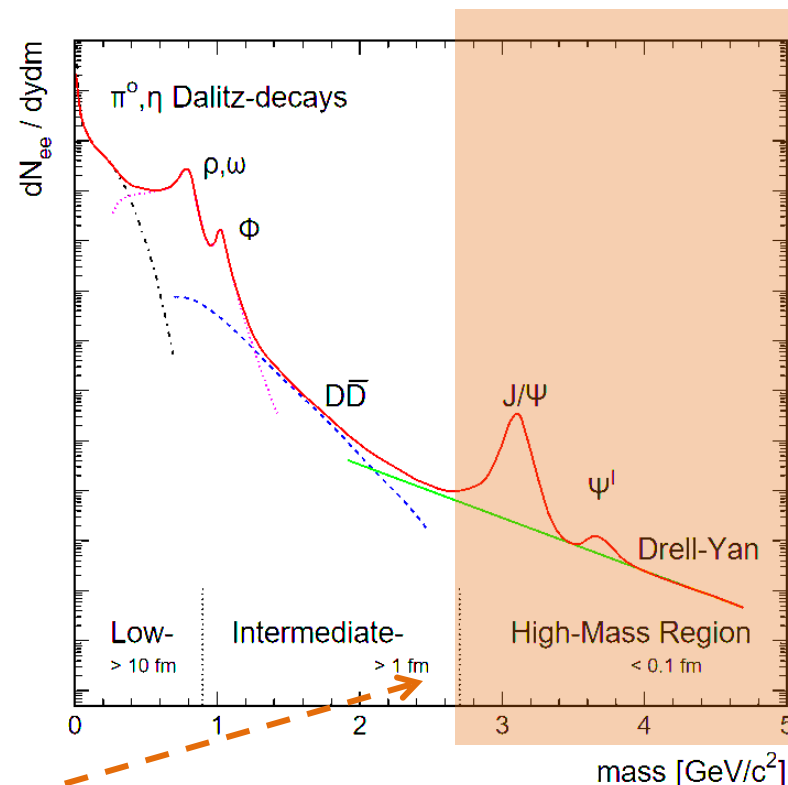The dielectron spectrum as a function of invariant mass in ultra-relativistic heavy-ion collisions [1].

o Motivation	o Analysis	o Results
• Heavy Ion Collisions	• Model: Random Forest Classifier	• ROC AUC Interpretation/ Precision, Recall and F-1 Scores
• Why Dielectrons? Why Machine Learning?	• Experimental Setup: Data/ Application of Model	o Conclusions

Why Dielectrons?

Electron-positron pairs *(dielectrons)* are the unique tools to study evolution of medium created in the collisions

- Due to lack of strong interactions they can be used to probe the inner regions of collisions
  - o no in medium effects on dielectrons
- They are produced at all stages of the collisions
  - o provide information about the whole space-time evolution of the system.

There is an approximate time ordering in the invariant mass of electron pairs:

- pairs with larger masses are produced early stage of the collision.
  - o Information about *initial state of the medium!!!*

The dielectron spectrum as a function of invariant mass in ultra-relativistic heavy-ion collisions [1].

o **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

o **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

o **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores

o **Conclusions**

Why Machine Learning?

A high purity sample of electron-positron pairs is required to measure the dielectron spectrum.

In dielectron analyses, various sources of background that are larger than pair signal by a few orders of magnitude have to be considered.

Rejection of those background components requires sophisticated analysis techniques.
• Traditional methods can provide high purity samples with **low signal efficiency:**
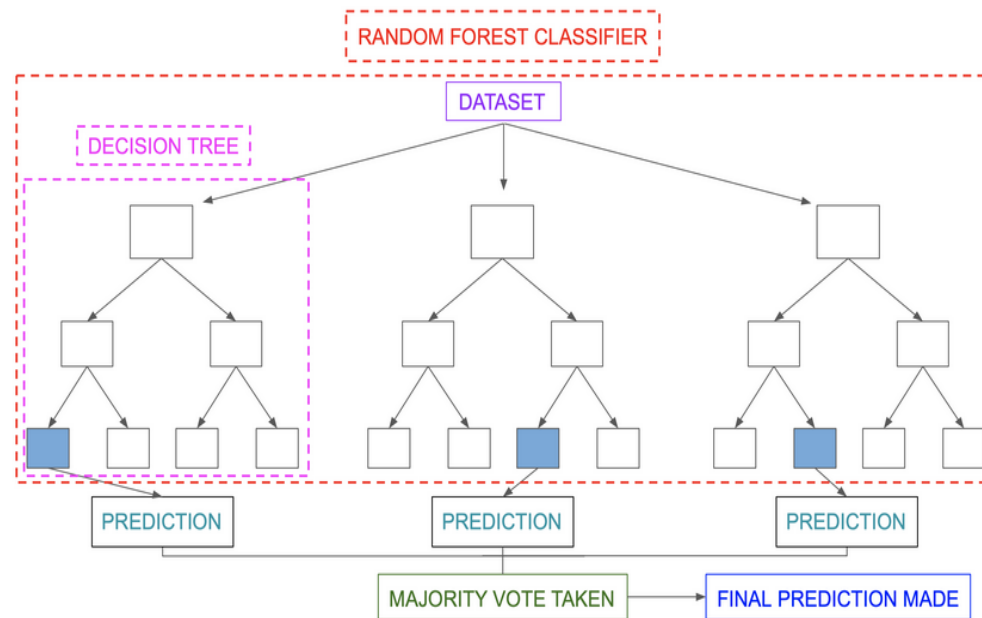 o QGP parameters can not be determined due to the high systematic uncertainties.

Artificial intelligence-based machine learning tools for pair identification could be used to improve dielectron spectrum with high efficiency.
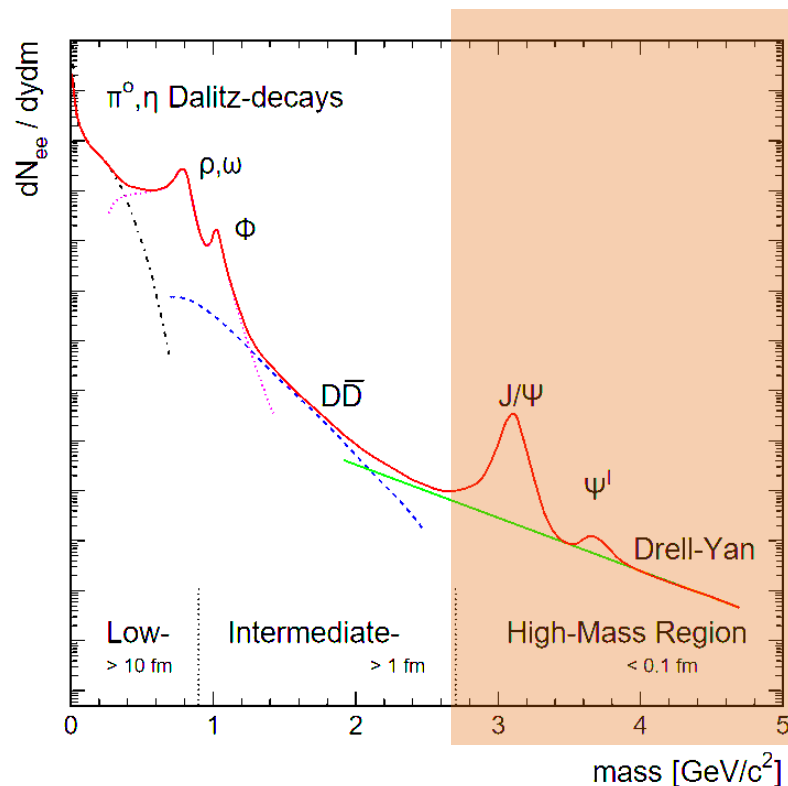
The dielectron spectrum as a function of invariant mass in ultra-relativistic heavy-ion collisions [1].

- ○ **Motivation**
  - • Heavy Ion Collisions
  - • Why Dielectrons? Why Machine Learning?
- ○ **Analysis**
  - • Model: Random Forest Classifier
  - • Experimental Setup: Data/ Application of Model
- ○ **Results**
  - • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
- ○ **Conclusions**

Model: Random Forest Classifier

- The model
  - ○ reproduces a set of Decision Trees
  - ○ regroups the votes from various Decision Trees to estimate the final class
- To make a prediction
  - ○ for regression: average the results
  - ○ for classification: majority vote

Schema of Random forest classifier algorithm [2].

o **Motivation**
- Heavy Ion Collisions
- Why Dielectrons? Why Machine Learning?

o **Analysis**
- Model: Random Forest Classifier
- Experimental Setup: Data/ Application of Model

o **Results**
- ROC AUC Interpretation/ Precision, Recall and F-1 Scores

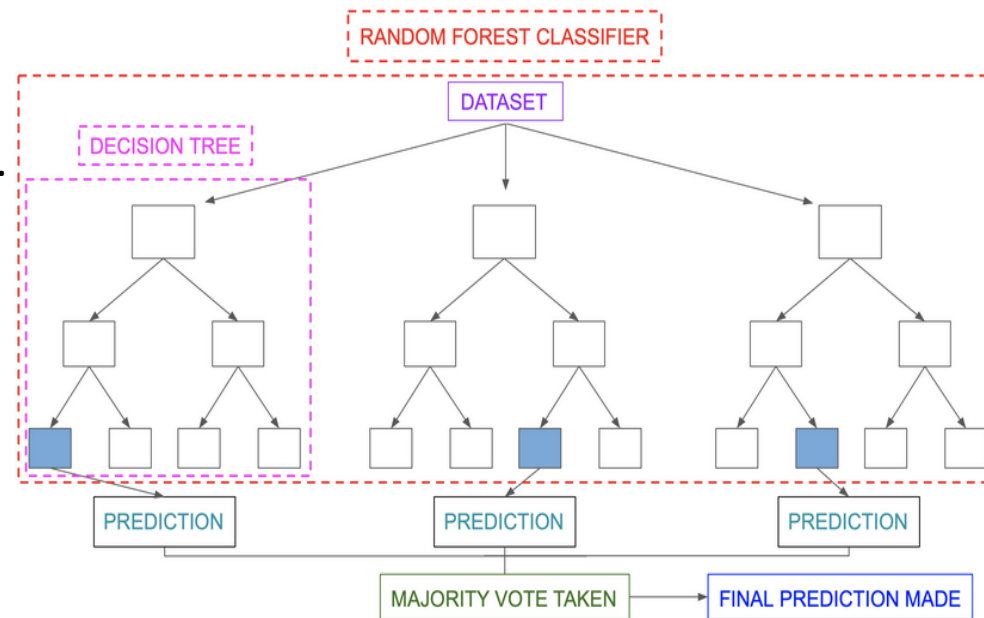o **Conclusions**

Model: Random Forest Classifier

- The model
  - o reproduces a set of Decision Trees
  - o regroups the votes from various Decision Trees to estimate the final class
- To make a prediction
  - o for regression: average the results
  - o for classification: majority vote

Advantages of the Model:
- Used for both classification and regression.
- Resistant to overfitting.
- Interpretable: Measure the relative importance of each feature on the prediction.
- More accurate compared to other algorithms.

Disadvantages of the Model:
- More resources are required for computation.
- Requires long time.

Schema of Random forest classifier algorithm [2].

o **Motivation**	o **Analysis**	o **Results**
• Heavy Ion Collisions	• Model: Random Forest Classifier	• ROC AUC Interpretation/ Precision, Recall and F-1 Scores
• Why Dielectrons? Why Machine Learning?	• Experimental Setup: Data/ Application of Model	o **Conclusions**

Experimental Setup: Data Set

LHC 2010 dielectron data set is used: 99912 pairs for inv. mass $2 - 110$ GeV/c^2

- o 56968 dielectron (e^-e^+) pairs : ***signal (57%)***
- o 42944 e^-e^- or e^+e^+ pairs: ***background (43%)***

  Signal% $\approx$ Background%

- Features used for pair classification:

High Level Features (HLF)
q_1, q_2 (charge)
η_1, η_2 (pseudorapidity)
ϕ_1, ϕ_2 (azimuthal angle)
p_1, p_2 (momentum)
M (invariant mass of pairs)
P (momentum of pairs)
θ (opening angle)

Inv. Mass Distribution of Dielectron Pairs from 2-110 GeV/c^2
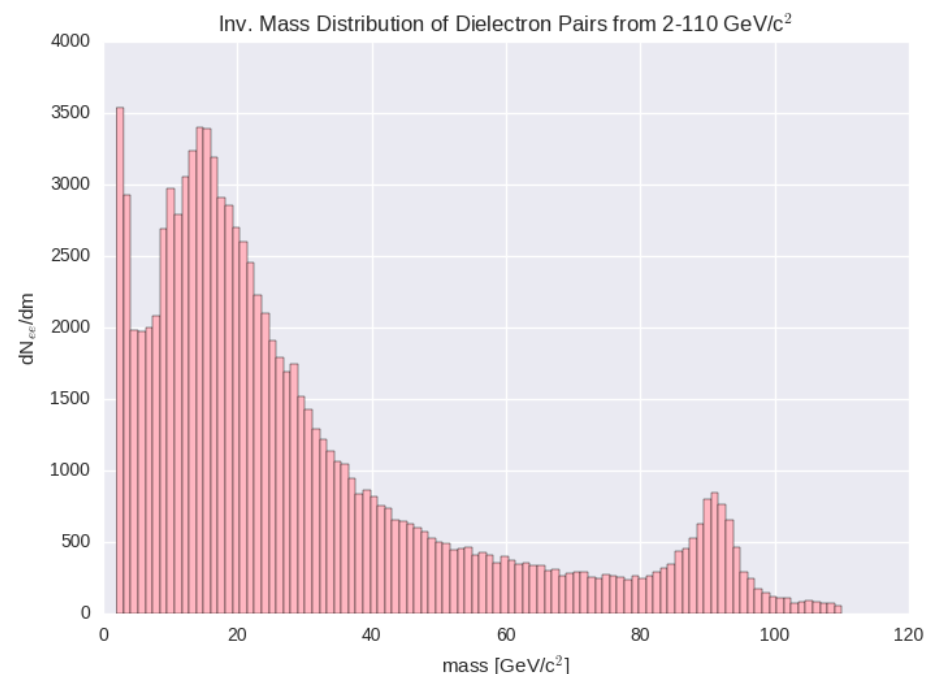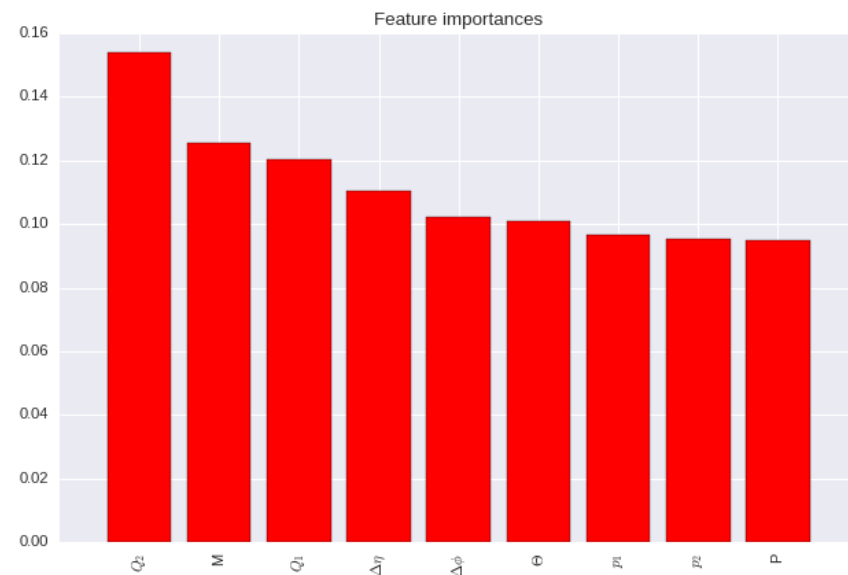
- Python implementation of the Random Forest classifier provided by scikit-learn package is used for identification of e^+e^- pairs.

o **Motivation**
   • Heavy Ion Collisions
   • Why Dielectrons? Why Machine Learning?

o **Analysis**
   • Model: Random Forest Classifier
   • Experimental Setup: Data/ Application of Model

o **Results**
   • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
o **Conclusions**

Experimental Setup: Application of Model

- To understand the impact of ML on classification of dielectron pairs , the classifier is trained by using HLF to see:
  o if HLF are good enough for distcriminate pairs.
  o if the highest importance feature matches with the used ones in traditional method.
  o if the pairs derived with the highest efficiency.

- Hyper parameters of the classifiers were tuned to have best classification.

- Train - test sample selection:
  o 60% Train and 40% Test

Feature importances

High Level Features (HLF)
q_1, q_2 (charge)
η_1, η_2 (pseudorapidity)
ϕ_1, ϕ_2 (azimuthal angle)
p_1, p_2 (momentum)
M (invariant mass of pairs)
P (momentum of pairs)
θ (opening angle)

○ **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

○ **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

○ **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
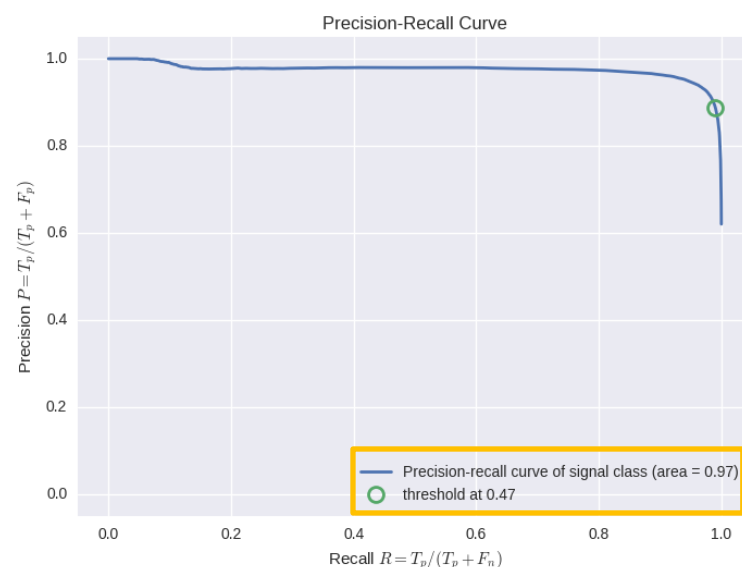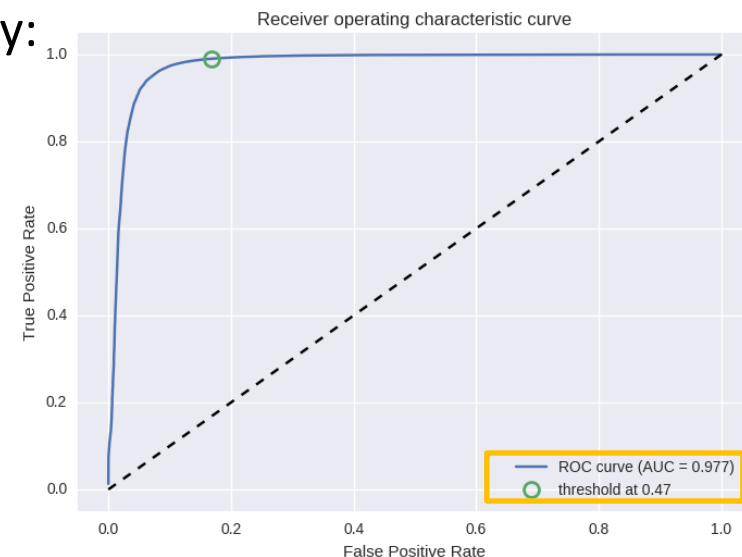○ **Conclusions**

Results: ROC AUC Interpretation

Pair identification efficiency and purity are studied by:

• Precision: $\frac{TP}{TP+FP}$ *"how accurate your signal predictions"*

• Recall: $\frac{TP}{TP+FN}$ *"how good your model to find all the actual signal"*

• F-1 Score: $\frac{2\times Precision \times Recall}{Precision + Recall}$

By using RFC with HLF to find e$^+$e$^-$ pairs:

  - The model has 97.7% chance to distinguish between signal and background.
  - The classifier reaches a recall of roughly 10% without any false positive predictions.

- o **Motivation**
  - • Heavy Ion Collisions
  - • Why Dielectrons? Why Machine Learning?
- o **Analysis**
  - • Model: Random Forest Classifier
  - • Experimental Setup: Data/ Application of Model
- o **Results**
  - • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
- o **Conclusions**

Results: ROC AUC Interpretation

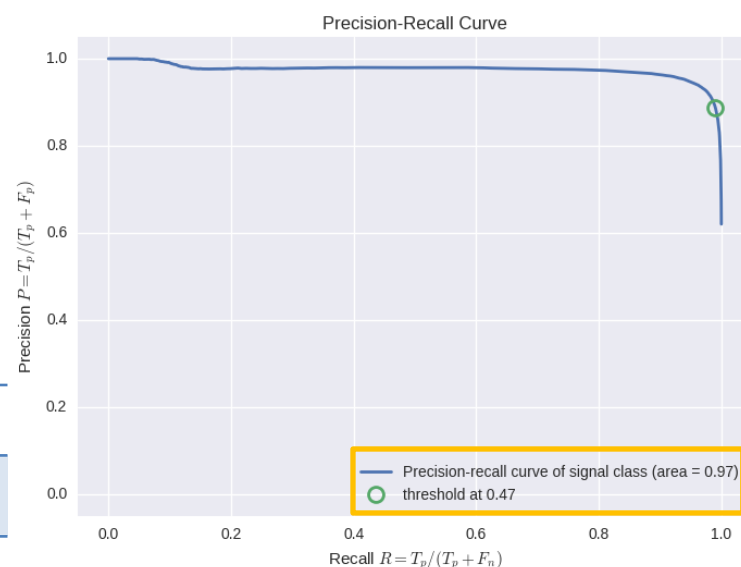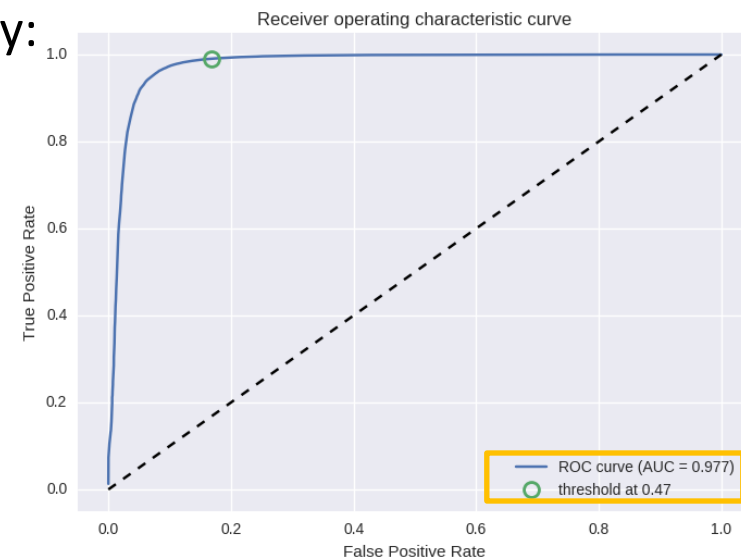Pair identification efficiency and purity are studied by:

- Precision: $\frac{TP}{TP+FP}$ *"how accurate your signal predictions"*

- Recall: $\frac{TP}{TP+FN}$ *"how good your model to find all the actual signal"*

- F-1 Score: $\frac{2 \times Precision \times Recall}{Precision + Recall}$

By using RFC with HLF to find e^+e^- pairs:

- The model has 97.7% chance to distinguish between signal and background.
- The classifier reaches a recall of roughly 10% without any false positive predictions.

RFC with HLF show precise and sensitive results.

Features	Avg. Precision	Avg. Recall	Avg. F-1 Score
HLF	0.93	0.92	0.92

Receiver operating characteristic curve

ROC curve (AUC = 0.977)
threshold at 0.47

Precision-Recall Curve

Precision-recall curve of signal class (area = 0.97)
threshold at 0.47

Recall $R = T_p/(T_p + F_n)$

o **Motivation**
  - Heavy Ion Collisions
  - Why Dielectrons? Why Machine Learning?

o **Analysis**
  - Model: Random Forest Classifier
  - Experimental Setup: Data/ Application of Model

o **Results**
  - ROC AUC Interpretation/ Precision, Recall and F-1 Scores
o **Conclusions**

Conclusions

Random Forest Classifier model is applied for e⁻e⁺ pair identification produced in high energy collisions to understand early stage of universe.

The results are showed:
- Selection of the features for signal classification in RF model has an important role on pair identification.
- Dielectrons are identified with Random Forest Classifier;
  - with higher precision and sensitivity.
  - by using the features used in traditional method.
- Without hard and time consuming background analysis the pairs can be identified with high efficiency.
- With Random Forest classifier has 97.7% chance to discriminate dielectron pairs in the right way.

Application of machine learning techniques is promising and may increase the quality of particle physics results.

Thank you!

This work is supported by TÜBİTAK-1001 119F302 and 2019TAEK(CERN)A5.H1.F5-23 projects.

○ **Motivation**
  • Heavy Ion Collisions
  • Why Dielectrons? Why Machine Learning?

○ **Analysis**
  • Model: Random Forest Classifier
  • Experimental Setup: Data/ Application of Model

○ **Results**
  • ROC AUC Interpretation/ Precision, Recall and F-1 Scores
○ **Conclusions**

References

[1] Rapp R., & Wambach J. (2002) In Advances in Nuclear Physics : Chiral symmetry restoration and dileptons in relativistic heavy-ion collisions.Springer.

[2] Mbaabu O. (2020, December 11). *Introduction to Random Forest in Machine Learning.* https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/